

A Statistical Framework for the Utilization of Simultaneous Pupil Plane and Focal Plane Telemetry for Exoplanet Imaging, Part I: Accounting for Aberrations in Multiple Planes

Richard A. Frazin

Dept. of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI 48109

ABSTRACT

A new generation of telescopes with mirror diameters of 20 m or more, called extremely large telescopes (ELTs) has the potential to provide unprecedented imaging and spectroscopy of exo-planetary systems, if the difficulties in achieving the extremely high dynamic range required to differentiate the planetary signal from the star can be overcome to a sufficient degree. Fully utilizing the potential of ELTs for exoplanet imaging will likely require simultaneous and self-consistent determination of both the planetary image and the unknown aberrations in multiple planes of the optical system, using statistical inference based on the wavefront sensor and science camera data streams. This approach promises to overcome the most important systematic errors inherent in the various schemes based on differential imaging, such as ADI and SDI. This paper is the first in a series on this subject, in which a formalism is established for the exoplanet imaging problem, setting the stage for the statistical inference methods to follow in the future. Every effort has been made to be rigorous and complete, so that validity of approximations to be made later can be assessed. Here, the polarimetric image is expressed in terms of aberrations in the various planes of a polarizing telescope with an adaptive optics system. Further, it is shown that current methods that utilize focal plane sensing to correct the speckle field, e.g., electric field conjugation, rely on the tacit assumption that aberrations on multiple optical surfaces can be represented as aberration on a single optical surface, ultimately limiting their potential effectiveness for ground-based astronomy.

Keywords: exoplanet, adaptive optics, short-exposure imaging, image processing

1. INTRODUCTION AND MOTIVATION

The coming decades will see the construction and operation of a new class of ground-based telescopes with mirror diameters of 20 m or more. Such instruments are called Extremely Large Telescopes (ELTs), and one of the top science priorities of the ELTs is direct imaging and spectroscopy of exoplanetary systems. ELTs should allow direct imaging of Earth-like planets,¹ but achieving this goal will require unprecedented precision in characterization of the optical system, as it is necessary to separate the starlight from the vastly fainter planetary emission. The most promising method for achieving high-contrast imaging and spectroscopy from the ground combines high-order adaptive optics (AO) with a stellar coronagraph. Typically, the AO system operates at visible wavelengths, while the science camera in the coronagraph captures images in the near-infrared. Observing in the near-IR improves the needed contrast ratios and is required for spectroscopy of key molecules such as water and methane. (For a review of high-contrast direct imaging, the reader may consult [2].) The major limitation in high-contrast imaging is the appearance of so-called "quasi-static" speckles (QSS) in the image that can be brighter than the planetary emission.^{3,4} These speckles are caused by constructive interference of the starlight, which has been distorted by aberrations in the optical system. QSS change on wide range of timescales due to a variety of mechanical stresses on the telescope caused by, for example, winds and temperature gradients, and they may last for days.

The current paradigm for treating the QSS is background subtraction based on differential imaging methods, the most important of which are angular differential imaging (ADI) and spectral differential imaging (SDI). ADI and SDI have number of biases due to complicated self-subtraction issues.^{5,6} The self-subtraction problems

E-mail: rfrazin *at* umich.edu

make these methods unsuitable for true imaging, and they create many challenges for characterizing point sources (planets), resulting in a large and ever-growing literature. As the basis of the SDI method is scaling of telescope point spread function (PSF) with wavelength, self-subtraction makes the method unsuited to imaging objects that are extended in the radial direction, and the method is sensitive to assumptions about the spectrum of the object. SDI can work well at large separations between the planet and star, but performs poorly at small separations since the wavelength-stretching effect is proportional to the distance from the center of the image. This is unfortunate because the expected success in planet discovery is heavily dependent on being able to make observations at small star-planet separations.^{7,8} ADI relies on the diurnal field rotation to separate the PSF from the planetary emission, so that any circularly symmetric component in the image will be indistinguishable from the PSF and be subject to self-subtraction. Combining SDI and ADI does not eliminate the self-subtraction problem, and resulting response pattern is rather complex.⁶ Further, SDI and ADI are plagued by a problem that the community has come to appreciate only very recently, which is that the uncertainties have been greatly under-estimated due to the fact that small sample statistics have not been taken into account. Briefly, the background-subtracted image has residual speckles, whose amplitudes decrease with distance from the star. In order to determine an uncertainty level as a function of radius, the amplitude of the speckles as a function of radial distance is estimated from the background-subtracted image. However, speckles have a size of $\sim \lambda/d$ (where λ is the wavelength and d is the telescope diameter), so, close to the host star there are very few speckles in a given small annulus. These small sample considerations lead to a statistical penalty that exponentially increases toward small separations.⁹

As the technology advances and higher levels of contrast are achieved, new problems associated with background subtraction via differential imaging are likely to surface. For example, if the star exhibits a small amount of linear polarization, the point spread matrix (PSM), which is the Stokes-vector generalization of the PSF [10], will depend on the diurnal rotation angle, further undermining the performance of ADI. Similar considerations apply to the planetary emission itself (which may be strongly linearly polarized due to Rayleigh scattering) if the PSM is not diagonally dominant. Furthermore, the PSFs of coronagraphic systems are sensitive to the pointing error, so that if the star is moving slightly relative to the pointing center during the diurnal rotation, the PSF will be changing, again, undermining the performance of ADI.

It is with this state of affairs in mind that several authors began to explore the utility of post-processing the pupil-plane telemetry, provided by a wavefront sensor (WFS), in conjunction with simultaneous image plane telemetry from a science camera, to treat the effects of the QSS.^{11–13} The historical context of this approach is reviewed in [11]. As the WFS must run at a frequency of approximately 1 kHz for the AO system to keep track of the atmospheric fluctuations, these methodologies also require ~ 1 kHz exposure cadence in the science camera, a prospect made practical by a new generation of ultra-low noise IR cameras capable of kHz readouts, such as the SWIR single photon detector, SAPHIRA eAPD and the MKIDS.^{14–17}

The combination of millisecond pupil plane and focal plane telemetry may well lead to orders of magnitude in contrast improvement over ADI and SDI, as it leverages a vastly larger and richer data set than standard exposure times, which average over the atmospheric turbulence. It is important to emphasize that the millisecond imaging techniques can be generalized to take advantage of essentially all constraints on problem proposed to date. These constraints include those imposed by diurnal rotation (used by ADI), multi-wavelength observations (used by SDI), as well polarization (used in polarization differential imaging [18]). Thus, the potential losses of information are rather limited. When one takes images that average over the atmospheric turbulence, the stellar speckles look much like planets. But, with an AO-corrected signal, planets and stellar speckles behave much differently at millisecond timescales. This is illustrated in Fig. 1, which shows a stellar coronagraph simulation result from [11]. Fig. 1 is a plot of two time series of the intensity calculated at a single pixel, corresponding to the location of a simulated planet, of the science camera. The dotted blue curve illustrates how the atmospheric turbulence (with AO correction) modulates the stellar speckle at a cadence of 1 millisecond. The solid black curve shows the much weaker modulation of the planetary light in the same pixel. These two time-series, the planetary intensity and the speckle intensity, are quite different in character, with the speckle having an approximately exponential probability density function (PDF), while the PDF of the planetary intensity is localized around its non-zero mean.^{11,19} This can be understood as follows: Much as that of the star, the planet's wavefront is stabilized by the AO system, as the flat part of the planet's wavefront is responsible for its intensity at this

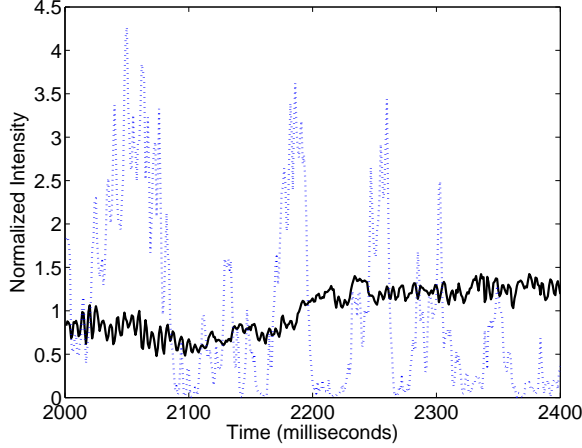


Figure 1. Modulation of intensities at a single pixel (corresponding to the planet location) of the science camera in a stellar coronagraph simulation. The black solid line shows the time-series of the temporal variation of the planetary intensity. The dotted blue line shows the corresponding stellar speckle intensity at the same pixel. Both the planetary and stellar intensity are normalized to have a mean of unity in this figure. The stellar intensity is enhanced by a sinusoidal static aberration at the spatial frequency corresponding to the planet's location. From [11], where this effect is demonstrated

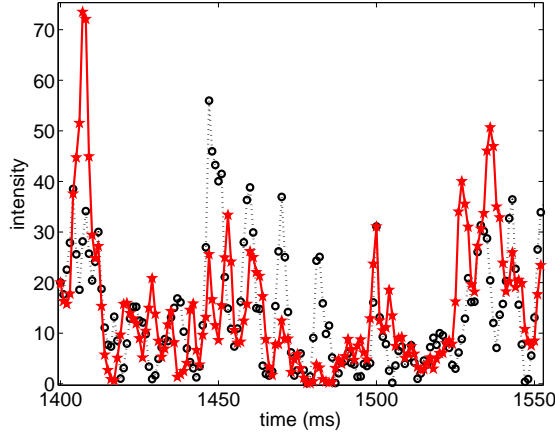


Figure 2. Stellar coronagraph simulation of the effect of a vibration on a speckle in the science camera. This speckle is caused by a vibrational NCPA of the form $\cos(\mathbf{k} \cdot \mathbf{r} - \omega t)$, and the modulation is caused by the interaction of the NCPA with the random AO residual. The black dotted curve corresponds to the intensity of a speckle for $\omega/2\pi = 100$ Hz and the red solid curve corresponds to $\omega/2\pi = 10$ Hz. The ms data are clearly sensitive to the frequency of the vibration. Both curves are calculated with the same sequence of residual phase values. From [13].

position in the image plane. However, the stellar speckle intensity at that location is entirely due to the random, non-flat part of the star's wavefront (the coronagraph removes the flat part) and, hence, it is much more volatile. In an actual measurement, the signal seen would be a weighted sum of these two curves, with the star having much more weight, necessitating the collection of many milliseconds of data and the utilization of statistical inference methods to separate them. Taking advantage of the fleeting moments in which the starlight is reduced at the planet's location, was first considered by Labeyrie [20].

The richness of the millisecond telemetry is further illustrated by Fig. 2, taken from [13], which shows its sensitivity to high-frequency vibrations. (Vibrations pose a particularly challenging problem to high-contrast astronomy.) Fig. 2 shows the time-dependence of the intensity of a vibrational speckle caused by a pupil-plane aberration of the form $\phi(\mathbf{r}) = \alpha \cos(\mathbf{k} \cdot \mathbf{r} + \omega t)$, where ω is the vibration frequency, t is the time, \mathbf{k} is the spatial frequency of the aberration and \mathbf{r} is the 2D spatial coordinate in the pupil plane. The black and red curves correspond to two different frequencies (10 and 100 Hz), and the fact that they do not coincide shows that the millisecond data are sensitive to the frequency. Obviously, conventional observations that use exposures orders of magnitude longer would see the same speckle, quite independently of frequency. Similarly, [13] showed that

when a pupil plane aberration is given by $\phi(r) = \alpha \cos(k \cdot r + \vartheta)$, the millisecond data are sensitive to whether α is real or imaginary as well as the phase angle ϑ , while conventional imaging cannot distinguish these effects.

In [11,13], the author's simulations use the millisecond telemetry from the focal and image planes to inform a large system of equations that provide a simultaneous statistical inference of both the aberrations in the optical system and the planetary image, avoiding background subtraction altogether. In contrast, in the background estimation method of [12], the millisecond telemetry is used to solve for the electric field due to the aberrations, as will be discussed later in Sec. 44.2. It is important to note that the methods of [11] and [12] are fully compatible with utilizing the same constraints that ADI and SDI, as well speckle nulling and electric field conjugation methods [21,22], which themselves will be discussed in Sec. 4.

The objective of this series of papers is to provide a rigorous statistical inference framework that is suitable for the high-precision requirements of exoplanet imaging with ELTs that are capable millisecond telemetry in both the pupil and image planes. The developments will account for polarization effects and detailed propagation within the optical system, noise and bandwidth effects in the wavefront sensor, as well as noise in the polarimetric image measured by the science camera.

The purpose of this paper, the first in the series, is to provide a formal description of polarimetric formation of an astronomical image in the presence of atmospheric turbulence, allowing for vector propagation of the wavefront through an adaptive optics telescope system with unknown aberrations at various surfaces in the optical system. The developments here also provide a new perspective on current methods that utilize focal plane sensing to treat the unwanted speckle field. Subsequent papers in this series will build upon this foundation to treat statistical inference of the aberrations and planetary image while treating practical limitations of the hardware.

2. ELEMENTARY CONCEPTS

The portion of the astronomical community with an interest in precise optical measurements, including those requiring ultra-high contrast, is increasingly becoming aware of the importance of polarization changes imparted to the light by the various elements of the telescope's optical system. These (often undesired) changes are generally called "polarization aberration," and are most commonly caused by reflection off mirrors at non-normal incidence. The effects of polarization aberration are clearly illustrated by the point-spread matrix in [10], which operates on the Stokes vector of the incoming light. This section introduces a straightforward formalism to treat polarimetric imaging system operating on a wave that is modulated by atmospheric turbulence. In addition, notations and conventions to be used throughout this series of papers are introduced here.

2.1 timescales and Coherence Functions

The developments presented below will involve three timescales. The smallest timescale is the coherence time τ_c of thermal light that has gone through a bandpass filter that removes all of the light except within a band of width $\Delta\nu$ centered on frequency ν , i.e., $\tau_c = 1/\Delta\nu$.²³ For example, setting the wavelength $\lambda = 2\mu\text{m}$ and a narrow bandpass of just 1%, $\tau_c \approx 6.7 \times 10^{-13}$ s. This timescale is relevant in the definition of the coherence functions, which require averaging over many periods of τ_c so that these functions obtain their mean values or very nearly so. The middle timescale, here referred to as the "Greenwood time," $\tau_G = 1/\nu_G$, where ν_G is the Greenwood frequency. The Greenwood time, on the order of 10^{-3} s, is the time period in which most of the atmospheric phase distortions change significantly and defines the bandwidth requirement for an effective AO system.²⁴ The largest timescale, τ_d is that for which the telescope optical systems exhibit dynamical aberration (due to time-variable mechanical stresses) that change the detailed structure of the image in the science camera. Precise values of these three timescales are of little importance in the developments below so long as they satisfy $\tau_c \ll \tau_G \ll \tau_d$, although the possibility of using millisecond observations to detect rapid mechanical vibrations, perhaps even with periods approaching 10^{-3} s is an intriguing one that is worth further investigation.¹³

Any spatially variant time-harmonic field can be represented in terms of an elliptically polarized wave with the electric field vector confined to some plane. However, in a field with arbitrary spatial variation, the normal vector specifying this plane will be a function of position, as will the polarization parameters.²⁵ In Sec. 33.2 it will be argued that atmospheric phase modulation will cause the normal vector describing the plane of polarization to wobble slightly on time-scale of τ_G . Rigorously accounting for this effect is beyond the scope of this paper and

is left to future work. Thus, it will be assumed that the electric field component of the light that the telescope is collecting can be represented as a quasi-monochromatic signal admitting a Jones vector representation in a known plane.^{25,26} Consider the following representation of the analytic signal [23, 25] corresponding to the amplitudes of the electric field fluctuations in orthogonal directions of a plane-wave propagating in the $+z$ direction :

$$\mathbf{E}(t) = \begin{bmatrix} E_x(t) \exp j[\Delta'_x(t) + \varsigma_x(t)] \\ E_y(t) \exp j[\Delta'_y(t) + \varsigma_y(t)] \end{bmatrix} \quad (1)$$

$$\equiv \begin{bmatrix} u_x(t) \\ u_y(t) \end{bmatrix} \equiv u(t), \quad (2)$$

which defines the Jones vector u to be used throughout this presentation, and where a factor of $\exp j[kz - 2\pi\nu t]$, with $k = 2\pi/\lambda$, has been suppressed. Constraints on time-dependent functions in Eq. (1) are described below. While $E_x(t)$ and $E_y(t)$ are in general complex-valued, $\Delta'_x(t)$, $\Delta'_y(t)$, $\varsigma_x(t)$ and $\varsigma_y(t)$ can be taken to be real. The random processes $\varsigma_x(t)$ and $\varsigma_y(t)$ are zero-mean so that $\langle \varsigma_x(t) \rangle_{\tau_c} = \langle \varsigma_y(t) \rangle_{\tau_c} = 0$, where the brackets $\langle \rangle_{\tau_c}$ indicate a time average in which, for some function $f(t)$:

$$\langle f(t) \rangle_{\tau_c} \equiv \frac{1}{\Delta t} \int_{t-\Delta t}^t dt' f(t'), \quad (3)$$

where $\tau_c \ll \Delta t \ll \tau_G$. Thus, $\langle f(t) \rangle_{\tau_c}$ is a low-pass version of $f(t)$.

The random processes $\varsigma_x(t)$ and $\varsigma_y(t)$ control the correlation between the x - and y - components, and they vary on timescales on the order of τ_c or longer, but much less than τ_G . The functions $E_x(t)$, $E_y(t)$, $\Delta'_x(t)$ and $\Delta'_y(t)$ are allowed to fluctuate on timescales on order of τ_G , but are quite constant over timescales similar to τ_c , so that, to a very high degree of accuracy, $E_x(t) = \langle E_x(t) \rangle_{\tau_c}$, $E_y(t) = \langle E_y(t) \rangle_{\tau_c}$, $\Delta'_x(t) = \langle \Delta'_x(t) \rangle_{\tau_c}$ and $\Delta'_y(t) = \langle \Delta'_y(t) \rangle_{\tau_c}$. In this way, the high-frequency character of the optical field has been placed into the $\exp(j2\pi\nu t)$, $\exp(j\varsigma_x(t))$ and $\exp(j\varsigma_y(t))$ factors. $E_x(t)$ and $E_y(t)$ are required to be in phase, so that

$$E_x(t) = |E_x(t)|e^{j\phi(t)} \quad \text{and} \quad E_y(t) = |E_y(t)|e^{j\phi(t)}, \quad (4)$$

which have the same phase factor $\phi(t)$. It will be useful to factor $u(t)$ into two parts, a scalar component representing the commonalities between u_x and u_y and a polarization state representing their differences:

$$u(t) = \underbrace{\sqrt{I(t)} \exp[j\phi(t)]}_{\text{scalar component}} \underbrace{\begin{bmatrix} e_x(t) \exp j[\Delta_x(t) + \varsigma_x(t)] \\ e_y(t) \exp j[\Delta_y(t) + \varsigma_y(t)] \end{bmatrix}}_{\text{polarization state}}, \quad (5)$$

where $I(t) = \sqrt{E_x(t)E_x^*(t) + E_y(t)E_y^*(t)}$ is the intensity, $0 \leq e_x(t), e_y(t) \leq 1$, $e_x^2(t) + e_y^2(t) = 1$, $\phi(t) = (\Delta'_x(t) + \Delta'_y(t))/2$, so that $\Delta_x(t) + \Delta_y(t) = 0$, and $*$ indicates complex conjugation. The form of Eq. (5) is convenient for easy interpretation. The scalar component controls the complex amplitude of the electric field vector, and in the polarization state, $e_x(t)$ and $e_y(t)$ set the relative amplitudes of the x - and y - components. $\Delta_x(t) - \Delta_y(t)$ determines linear vs. circular polarization. The degree of polarization is set by $e_x(t)$, $e_y(t)$, $\varsigma_x(t)$ and $\varsigma_y(t)$.

For an astronomical signal just above the Earth's atmosphere, I , e_x , e_y , ϕ , Δ_x and Δ_y are essentially constant in time, so the factorization in Eq. (5) becomes:

$$\begin{aligned} u(t) &= \sqrt{I} \exp[j\phi] \begin{bmatrix} e_x \exp j[\Delta_x + \varsigma_x(t)] \\ e_y \exp j[\Delta_y + \varsigma_y(t)] \end{bmatrix} \\ &\equiv \sqrt{I} \exp[j\phi] \check{u}(t), \end{aligned} \quad (6)$$

thus defining the polarization state vector $\check{u}(t)$. The time-dependent effects of the atmosphere and the telescope system will be assumed to be representable as linear operators, which take the form of a Jones matrix, as described below in Sec. 2.2.2. The fluctuations related to atmospheric and telescopic dynamics on the timescales of τ_G and τ_d , respectively, do not cause depolarization, but they can cause other changes in the polarization.

When an astronomical signal of the form in Eq. (6) is multiplied by a time-dependent Jones matrix that varies on timescales of τ_G or longer, the resulting field can be described by the more general form in Eq. (5).

Describing the intensities measured by polarization sensitive instruments requires the 2nd order statistics of u , which are contained in the 4×1 coherency vector (usually called the coherency matrix when arranged in the form of a 2×2 matrix),^{23, 25} and is given by the time-averaged Kronecker product of u and u^* , denoted as $u \otimes u^*$ [27]:

$$J(t) = \langle u(t) \otimes u^*(t) \rangle_{\tau_c} \\ = I(t) \exp[-2\Im(\phi(t))] \begin{bmatrix} e_x^2(t) \\ e_x(t)e_y(t) \exp[j\Delta_x(t) - \Delta_y(t)] \langle \exp j[\varsigma_x(t) - \varsigma_y(t)] \rangle_{\tau_c} \\ e_y(t)e_x(t) \exp[j\Delta_y(t) - \Delta_x(t)] \langle \exp j[\varsigma_y(t) - \varsigma_x(t)] \rangle_{\tau_c} \\ e_y^2(t) \end{bmatrix}. \quad (7)$$

where the \Im symbol indicates the imaginary part. The second equality in Eq. (7) makes explicit use timescale separations explained above. In order that the linear vs. circular polarization nature of the light is set only by $\Delta_x(t) - \Delta_y(t)$, it is clear that the condition $\langle \varsigma_x(t) - \varsigma_y(t) \rangle_{\tau_c} = 0$ must be satisfied, as any non-zero mean would modify the relative phase of the x - and y -components of the field.

Most intensity measurements are more readily interpreted in terms of the Stokes vector, which is easily obtained from J :

$$J^S(t) = \mathbf{Q}J(t), \text{ where } \mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 1 & 0 \\ 0 & -j & j & 0 \end{bmatrix}, \quad (8)$$

where the 0th component of J^S is the total intensity, the 1st corresponds to vertical/horizontal polarization, the 2nd to diagonal polarization and the 3rd to circular.^{25, 26} Note that \mathbf{Q} is invertible and has 4 eigenvalues that all have an absolute value of $\sqrt{2}$. Thus, the coherency and Stokes vectors are equivalent representations, and it follows that a measurement of the full Stokes vector fully specifies the beam, allowing determination of $I(t) \exp[-2\Im(\phi(t))]$, $e_x(t)$, $e_y(t)$, $\Delta_x(t) - \Delta_y(t)$, and $\langle \exp j[\varsigma_x(t) - \varsigma_y(t)] \rangle_{\tau_c}$.

When considering wave propagation, it is necessary to allow the field to have spatial dependence in the transverse plane. Let \mathbf{r}_1 and \mathbf{r}_2 be the transverse [i.e., (x, y)] coordinates of two points in some plane, and let t_1 and t_2 be two times. The 4×1 cross-density vector, $\Gamma(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2)$, is a generalization of the coherency ([28] calls this the "cross-spectral density"). In the most general case, the 4×1 cross-density vector is defined as $\Gamma(\mathbf{r}_1, \mathbf{r}_2, t_1, t_2) \equiv \langle u(\mathbf{r}_1, t_1) \otimes u^*(\mathbf{r}_2, t_2) \rangle_{\tau_c}$, so that $\Gamma(\mathbf{r}, \mathbf{r}, t, t) = J(\mathbf{r}, t)$. For the purposes of image formation it is sufficient to only allow time differences $\tau \equiv t_2 - t_1$ such that $|\tau| < \tau_c$. As the ς_x and ς_y stochastic processes were assumed to be stationary, the rapidly fluctuating part of the cross-density (i.e., at timescales $\ll \tau_G$) can only be a function of the time difference τ .²³ Further, as the fields are taken to be quasi-monochromatic, $\Gamma(\mathbf{r}_1, \mathbf{r}_2, t, t + \tau) = \langle u(\mathbf{r}_1, t) \otimes u^*(\mathbf{r}_2, t + \tau) \rangle_{\tau_c} = \langle u(\mathbf{r}_1, t) \otimes u^*(\mathbf{r}_2, t) \rangle_{\tau_c} \exp(-j2\pi\nu\tau)$. The mutual coherency γ is defined as the cross-density evaluated at 0 time difference, so, $\gamma(\mathbf{r}_1, \mathbf{r}_2, t) \equiv \Gamma(\mathbf{r}_1, \mathbf{r}_2, t, t) = \langle u(\mathbf{r}_1, t) \otimes u^*(\mathbf{r}_2, t) \rangle_{\tau_c}$. As will be seen in Sec. 22.2, the mutual coherency plays an important role in polarimetric image formation.

Consider the field of an astronomical source, as represented in Eq. (6), at the spatial location \mathbf{r} in some plane just above the Earth's atmosphere. Clearly, I , e_x , e_y , Δ_x and Δ_y must be independent of \mathbf{r} because the Stokes parameters that would be measured in that plane (at least not too far from Earth) are independent of \mathbf{r} . But, in Sec. 33.1 it will be seen that the mutual coherency $\gamma(\mathbf{r}_1, \mathbf{r}_2)$ carries an encoding of the image of the astronomical source. Therefore, the image information must be contained the functions ϕ , ς_x and ς_y , which must be functions of the spatial coordinate \mathbf{r} . To see this, consider the definition of the mutual coherency. Similarly to Eq. (7),

the mutual coherency is given by:

$$\begin{aligned}\gamma(\mathbf{r}_1, \mathbf{r}_2) &= \langle u(\mathbf{r}_1, t) \otimes u^*(\mathbf{r}_2, t) \rangle_{\tau_c} \\ &= I \exp j[\phi(\mathbf{r}_1) - \phi(\mathbf{r}_2)] \times \\ &\quad \begin{bmatrix} e_x^2 \langle \exp j[\zeta_x(\mathbf{r}_1, t) - \zeta_x(\mathbf{r}_2, t)] \rangle_{\tau_c} \\ e_x e_y \exp j[\Delta_x - \Delta_y] \langle \exp j[\zeta_x(\mathbf{r}_1, t) - \zeta_y(\mathbf{r}_2, t)] \rangle_{\tau_c} \\ e_y e_x \exp j[\Delta_y - \Delta_x] \langle \exp j[\zeta_y(\mathbf{r}_1, t) - \zeta_x(\mathbf{r}_2, t)] \rangle_{\tau_c} \\ e_y^2 \langle \exp j[\zeta_y(\mathbf{r}_1, t) - \zeta_y(\mathbf{r}_2, t)] \rangle_{\tau_c} \end{bmatrix},\end{aligned}\quad (9)$$

where the time argument of γ has been dropped because it does not depend on time for astronomical signals above the atmosphere. As $\phi(\mathbf{r})$ is strictly real above the atmosphere, it has no influence on the amplitude of $\gamma(\mathbf{r}_1, \mathbf{r}_2)$. Thus, the amplitude of $\gamma(\mathbf{r}_1, \mathbf{r}_2)$ is determined by the function $\langle \exp j[\zeta_x(\mathbf{r}_1, t) - \zeta_y(\mathbf{r}_2, t)] \rangle_{\tau_c}$ and its three cousins in Eq. (9).

According to the Van Cittert-Zernike theorem for scalar fields, the astronomical image can be determined by taking the Fourier transform of the mutual coherence function and [23]. Since vector analog of the mutual coherence function is the mutual coherency in Eq. (9), the polarimetric image of the astronomical source is encoded in $\phi(\mathbf{r}_1) - \phi(\mathbf{r}_2)$ (which is analogous to the phase of the mutual coherence function), and the functions $\langle \exp j[\zeta_x(\mathbf{r}_1, t) - \zeta_x(\mathbf{r}_2, t)] \rangle_{\tau_c}$, $\langle \exp j[\zeta_x(\mathbf{r}_1, t) - \zeta_y(\mathbf{r}_2, t)] \rangle_{\tau_c}$ and $\langle \exp j[\zeta_y(\mathbf{r}_1, t) - \zeta_y(\mathbf{r}_2, t)] \rangle_{\tau_c}$ (which are analogous to the magnitude of the mutual coherence function, also known as the "visibility" in long-baseline stellar interferometry). This will be discussed further in Sec. 33.1. As discussed after Eq. (7), $\zeta_x(\mathbf{r}, t)$ and $\zeta_y(\mathbf{r}, t)$ must satisfy the condition $\langle \zeta_x(\mathbf{r}, t) - \zeta_y(\mathbf{r}, t) \rangle_{\tau_c} = 0$, but the value of $\langle \zeta_x(\mathbf{r}_1, t) - \zeta_y(\mathbf{r}_2, t) \rangle_{\tau_c}$ has no such restriction on its behavior (and similarly for its x, x and y, y cousins). To understand this, consider an extended astronomical source with purely linear 45° polarization, so ($e_x = e_y = 1/\sqrt{2}$), i.e., the coherency of the light emerging from the source is given by $S(\boldsymbol{\alpha}) = s(\boldsymbol{\alpha})S$, in which $\boldsymbol{\alpha}$ corresponds to the position in the sky, $s(\boldsymbol{\alpha})$ is a real-valued scalar function, and the vector $S = \mathbf{Q}^{-1}[1, 0, 1, 0]^T$ (where T indicates transpose). In this case, the source has a spatially constant polarization, so the mutual coherency seen at the Earth must exhibit scalar behavior, i.e., $\gamma(\mathbf{r}_1, \mathbf{r}_2) = g(\mathbf{r}_2 - \mathbf{r}_1)J$, where J is a constant coherency vector and $g(\mathbf{r})$ is some scalar function. The scalar behavior of γ requires $\langle \zeta_x(\mathbf{r}_1, t) - \zeta_x(\mathbf{r}_2, t) \rangle_{\tau_c} = \langle \zeta_x(\mathbf{r}_1, t) - \zeta_y(\mathbf{r}_2, t) \rangle_{\tau_c} = \langle \zeta_y(\mathbf{r}_1, t) - \zeta_y(\mathbf{r}_2, t) \rangle_{\tau_c} = 0$. Further, one has $\langle \exp j[\zeta_x(\mathbf{r}_1, t) - \zeta_x(\mathbf{r}_2, t)] \rangle_{\tau_c} = \langle \exp j[\zeta_x(\mathbf{r}_1, t) - \zeta_y(\mathbf{r}_2, t)] \rangle_{\tau_c} = \langle \exp j[\zeta_y(\mathbf{r}_1, t) - \zeta_x(\mathbf{r}_2, t)] \rangle_{\tau_c} = \langle \exp j[\zeta_y(\mathbf{r}_1, t) - \zeta_y(\mathbf{r}_2, t)] \rangle_{\tau_c}$. However, for realistic sources in which the polarization state is not spatially constant, these conditions do not hold, allowing the "polarization ellipse," if it were ever defined for the mutual coherency, to change as a function of $\mathbf{r}_1 - \mathbf{r}_2$.

2.2 Polarimetric Image Formation and Polarization Aberration

One may model the effect of a perfectly flat optical surface, such as an ideal flat mirror, on a plane wave with a 2×2 Jones matrix Υ operating on the Jones vector u , so that the outgoing plane wave is given by $u' = \Upsilon u$.²⁶ The coherency vector of the outgoing wave is given by $J' = (\Upsilon u) \otimes (\Upsilon^* u^*) = (\Upsilon \otimes \Upsilon^*)(u \otimes u^*) = (\Upsilon \otimes \Upsilon^*)J$, which makes use of the common identity for Kronecker products of matrices: $AC \otimes BD = (A \otimes B)(C \otimes D)$.²⁹ As explained in the fine paper by McGuire & Chipman [27], so long as the paraxial approximation applies, this result can be extended to any optical system that is non-scattering and does not depolarize the light.¹⁰ Let the field in entrance pupil plane, denoted with index 0, of some such optical system be represented by the Jones vector $u_0(\mathbf{r}_0, t)$, where \mathbf{r}_0 is the 2D coordinate vector in the plane, and let the outgoing field in the exit plane (with index number 1) be represented by $u_1(\mathbf{r}_1, t)$. For example, plane 0 could correspond to the entrance pupil of a telescope and the plane 1 could correspond to the surface of a detector where an image is formed. The optical system is modeled in terms of a Jones propagation matrix $\Upsilon(\mathbf{r}_1, \mathbf{r}_0)$:

$$u_1(\mathbf{r}_1, t) = \int d\mathbf{r}_0 \Upsilon(\mathbf{r}_1, \mathbf{r}_0) u_0(\mathbf{r}_0, t), \quad (10)$$

where $\Upsilon(\mathbf{r}_1, \mathbf{r}_0)$ is 0 when \mathbf{r}_0 is not within the spatial limits of the entrance pupil. At first glance, Eq. (10) implies instantaneous action and violates the laws of relativity, however, the time delays in the optical system are accounted for by phase shifts, such as a quadratic phase term describing a lens or focusing mirror.³⁰ As per

the conventions *mise en place* in the Appendix, the t in Eq. (10) only refers to fluctuations on timescales far longer than τ_c and the action can be considered to be instantaneous at such timescales. The coherency vector of the outgoing light is given by:

$$J_1(\mathbf{r}_1, t) = u_1(\mathbf{r}_1, t) \otimes u_1^*(\mathbf{r}_1, t) = \int d\mathbf{r}_0 \int d\mathbf{r}'_0 [\Upsilon(\mathbf{r}_1, \mathbf{r}_0) \otimes \Upsilon^*(\mathbf{r}_1, \mathbf{r}'_0)] \gamma_0(\mathbf{r}_0, \mathbf{r}'_0, t), \quad (11)$$

where γ_0 is the mutual coherency in plane 0. The 4×4 matrix of functions $\Upsilon(\mathbf{r}_1, \mathbf{r}_0) \otimes \Upsilon^*(\mathbf{r}_1, \mathbf{r}'_0)$ is the propagation kernel for the mutual coherency. Thus, the coherency vector output by the optical system can be expressed in terms of the Jones propagator of the optical system and the mutual coherency vector at the entrance pupil.¹⁰

Below [in Eq. (14)], it will be seen that γ is essentially the spatial Fourier transform of the scene to be imaged. Therefore, if Υ in Eq. (11) has Fourier transforming properties, then $J_1(\mathbf{r}, t)$ will be a polarimetric image of the scene. As is emphasized in [10], only an optical system in which $\Upsilon \otimes \Upsilon$ is proportional to the identity matrix will treat all Stokes parameters of the input beam the same way, and any deviation from this ideal is called polarization aberration. Examples of polarization aberration in telescopes are discussed in [10], and [27] analyzes a circular retarding lens (with corn syrup as the key ingredient) for which left and right circularly polarized light have different focal lengths.

3. PROPAGATION MODELS

Eq. (11) relates the polarimetric image produced by an imaging system to the mutual coherency of the optical radiation incident upon its entrance pupil. The objective of this section to follow the mutual coherency of the light starting from the astronomical object to the Earth, through the atmosphere and, finally, through the telescope while accounting for unknown aberrations and the vector nature of the field. The derivation given in Sec. 33.1 for the polarimetric version of the van Cittert-Zernike theorem is similar to the scalar version in [23]. Henceforth, this paper will use the notation conventions given in the Appendix.

3.1 To Earth

Consider a planetary system, as viewed from the direction of Earth, and assume that the telescope is pointed somewhere close to the center of the system, where the host star is located. The pointing direction of the telescope corresponds to the $-z$ direction, so that light from the planetary system collected by the telescope is traveling in more-or-less the $+z$ direction. Then, any point in the planetary system is given by the coordinates $(z_0\boldsymbol{\alpha}, -z_0)$, where z_0 is the distance from the planetary system to Earth, and $\boldsymbol{\alpha}$ is a two-dimensional vector, with units of radians, corresponding to the direction cosines $-x/z_0$ and $-y/z_0$.

At optical frequencies, outer space is effectively homogenous and isotropic, so there are no polarization effects and wave propagation is a scalar phenomenon. Then, the Huygens-Fresnel principle can be used to propagate each component of u independently to a plane (normal direction z) just above the Earth's atmosphere.³⁰ Denote this plane with the index -1 (below, the 0 plane will correspond to the telescope entrance pupil, and the positive indices will correspond to subsequent planes within the optical system). The optical radiation emerging from the planetary system can be considered to be emitted from the $z = -z_0$ plane, which, say, cuts through the star, and the field emerging from the plane is given by $u_p(z_0\boldsymbol{\alpha})$. Then, the planetary system gives rise to the vector field $u_{p-1}(\mathbf{r})$ at location \mathbf{r} in the -1 plane, just above Earth's atmosphere, which is given by [23]:

$$u_{p-1}(\mathbf{r}) = \frac{-k}{2\pi} \int_p d(z_0^2\boldsymbol{\alpha}) u_p(z_0\boldsymbol{\alpha}) \frac{\exp[-jkr(\mathbf{r}, \boldsymbol{\alpha})]}{r(\mathbf{r}, \boldsymbol{\alpha})} \cos\vartheta(\mathbf{r}, \boldsymbol{\alpha}), \quad (12)$$

where u_p is the field emerging from the planetary system, $r^2(\mathbf{r}, \boldsymbol{\alpha}) = z_0^2 + (\mathbf{r} - z_0\boldsymbol{\alpha})^2$, and $\cos\vartheta(\boldsymbol{\alpha}, \mathbf{r})$ is the obliquity factor.

As the planetary system consists of many independently radiating sources, the emergent light is effectively spatially incoherent, so that the mutual coherency of the emergent radiation is given by $u_p(z_0\boldsymbol{\alpha}_1) \otimes u_p^*(z_0\boldsymbol{\alpha}_2) =$

$(4\pi/k^2)J_p(z_0\boldsymbol{\alpha}_1)\delta(\boldsymbol{\alpha}_2 - \boldsymbol{\alpha}_1)$.^{*} Using the spatially incoherent source condition and the fact that $|z_0| \gg |\mathbf{r}|$, one finds:

$$\begin{aligned}\gamma_{p-1}(\mathbf{r}_1, \mathbf{r}_2) &= u_{p-1}(\mathbf{r}_1) \otimes u_{p-1}^*(\mathbf{r}_2) \\ &= \frac{\exp[jp(\mathbf{r}_1, \mathbf{r}_2)]}{\pi} \int_p d\boldsymbol{\alpha} J_p(z_0\boldsymbol{\alpha}) \exp[jk(\mathbf{r}_1 - \mathbf{r}_2) \cdot \boldsymbol{\alpha}],\end{aligned}\quad (13)$$

where the quadratic phase factor is $p(\mathbf{r}_1, \mathbf{r}_2) = (k/2z_0)(\mathbf{r}_1 \cdot \mathbf{r}_1 - \mathbf{r}_2 \cdot \mathbf{r}_2)$. For astronomical sources the $\exp[jp(\mathbf{r}_1, \mathbf{r}_2)]$ factor can be dropped since the telescope diameter squared is much smaller than $2z_0/k$ (e.g., for a 100 m telescope operating at 1 μm and targeting *Alpha Centauri A*, this ratio is at most $\approx 10^{-6}$).

Note that in classical radiometry $J_p(z_0\boldsymbol{\alpha})$ has units of [energy/(area solid-angle time)] and is called the radiance, and $d\boldsymbol{\alpha}$ has units of [solid-angle].²⁵ In astronomy, the preference is for surface brightness, which also has units [energy/(area solid-angle time)], but is a function of the angular coordinate $\boldsymbol{\alpha}$ instead of the linear coordinate $(z_0\boldsymbol{\alpha})$.³¹ Define the planetary surface brightness coherency vector as $S(\boldsymbol{\alpha}) \equiv J_p(z_0\boldsymbol{\alpha})$ [note that this does not constitute a change of variables, so scaling by the Jacobian is not required], so that Eq. (13) becomes:

$$\gamma_{p-1}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{\pi} \int_p d\boldsymbol{\alpha} S(\boldsymbol{\alpha}) \exp[jk(\mathbf{r}_1 - \mathbf{r}_2) \cdot \boldsymbol{\alpha}]. \quad (14)$$

Eq. (14) is the vector-field generalization of the well-known van Cittert-Zernike theorem for astronomical sources, and states that the mutual coherency vector of the light arriving at the Earth is proportional to the Fourier transform of the coherency vector of the light emerging from the planetary system, $S(\boldsymbol{\alpha})$. The objective of this paper is to develop a methodology for determining $S(\boldsymbol{\alpha})$ in the high-contrast context. Tervo et al. [32] derive the same result in terms of the Stokes parameters and discuss the history of previous similar efforts.

The coherency of the planetary light above the atmosphere is easily calculated from Eq. (14) as $J_{p-1}(\mathbf{r}) = \gamma_{p-1}(\mathbf{r}, \mathbf{r})$, resulting in:

$$J_p = \frac{1}{\pi} \int_p d\boldsymbol{\alpha} S(\boldsymbol{\alpha}). \quad (15)$$

Note that J_p has no dependence on the spatial coordinate \mathbf{r} .

In principle, Eq. (14) describes the mutual coherency arising from the entire planetary system, central star included. However, in the high-contrast context the star is vastly brighter than the surrounding material and the propagation of its light must be treated with much more care than the planetary light. Therefore, the starlight and planetary light are to be treated separately in these developments. Assume that the star is unresolved, essentially acting as a point source, and is located at a small angle $\boldsymbol{\alpha}_\star$ from the telescope pointing direction, and let J_\star be its coherency vector, as would be measured above the atmosphere. The star's mutual coherency can be calculated with the aid of Eq. (14) by setting $S(\boldsymbol{\alpha}) = \pi J_\star \delta(\boldsymbol{\alpha} - \boldsymbol{\alpha}_\star)$, resulting in:

$$\gamma_{\star-1}(\mathbf{r}_1, \mathbf{r}_2) = J_\star \exp[jk(\mathbf{r}_1 - \mathbf{r}_2) \cdot \boldsymbol{\alpha}_\star]. \quad (16)$$

Since $\gamma_{\star-1}(\mathbf{r}_1, \mathbf{r}_2) = u_{\star-1}(\mathbf{r}_1) \otimes u_{\star-1}^*(\mathbf{r}_2)$, one can see that $u_{\star-1}$ can be factored as per Eq. (6):

$$u_{\star-1}(\mathbf{r}) = \sqrt{I_\star} \exp(jk\boldsymbol{\alpha}_\star \cdot \mathbf{r}) \check{u}_\star \quad (17)$$

where I_\star is star's intensity, $k\boldsymbol{\alpha}_\star \cdot \mathbf{r}$, is the real-valued phase, and \check{u}_\star is the polarization state vector. With the variables so-defined, one has:

$$\begin{aligned}J_\star &= \gamma_{\star-1}(\mathbf{r}, \mathbf{r}) = u_{\star-1}(\mathbf{r}) \otimes u_{\star-1}^*(\mathbf{r}) \\ &= I_\star \check{u}_\star \otimes \check{u}_\star^*.\end{aligned}\quad (18)$$

^{*} The only truly spatially incoherent source corresponds to $\gamma = 0$, and [23] advocates approximating the incoherent condition with the scale factor used here, corresponding to a field that is coherent over an area roughly the size of λ^2 .

3.2 Through the Atmosphere

The journey from the -1 plane (above the atmosphere) to the 0 plane, corresponding to the entrance pupil of the telescope, requires the light to traverse the Earth's atmosphere. While atmospheric polarization aberration has been discussed in the military context of propagation of beams of laser light ([33] has a useful list of references) the literature on the subject in the context of astronomy appears to be rather limited. The most important effect of the atmosphere is that of a random phase screen, and the second most important effect is scintillation, which itself is caused by phase screening at larger heights.³⁴ Absorption effects by the atmosphere should only affect photometry and are generally well-understood.³¹ The atmosphere generally not expected to produce polarization effects, however, several studies have found polarization related to presences of specific particles or molecules or molecules in the atmosphere. For example, [?] found small, wavelength-dependent linear polarization in high-precision solar observations at large zenith angles, and attributed it to double scattering off of various molecules. Bailey et al. [35] report a slight linear polarization caused by Saharan dust high in the atmosphere over the Canary Islands and cite previous studies indicating polarization effects on the order of 1 part in 10^6 . There is also literature on atmospheric polarization effects in the context of laser beam propagation for military purposes. For example, [36] gives results for atmospheric polarization effects on a Gaussian Schell beam. While it seems that the case of a plane wave incident upon the Earth from an astronomical source should be representable by a Gaussian Schell beam in the limit that the beam-width parameters go to infinity, the laser propagation results given by [36] are unphysical in that limit.

Given the current state of knowledge of atmospheric polarization effects, it will be assumed that the atmosphere contributes no polarization aberration, so the star's Stokes parameters are known and correspond to those that can be measured in a standard observational setting (with long exposures that average over the turbulence), which is likely to be correct to first order. Thus, the effect of the atmospheric turbulence will be considered to be a scalar phenomenon described by a complex-valued phase screen (imaginary values account for scintillation) $\phi_a(\mathbf{r}, t)$. Then, as there are no optics between the -1 plane (above the atmosphere) and the 0 plane (the telescope entrance pupil), the Jones vector of the starlight at the entrance pupil is given by:

$$\begin{aligned} u_{*0}(\mathbf{r}, t) &= u_{*-1}(\mathbf{r}) \exp[j\phi_a(\mathbf{r}, t)] \\ &= \sqrt{I_*} \exp[j[k\boldsymbol{\alpha}_* \cdot \mathbf{r} + \phi_a(\mathbf{r}, t)]] \check{u}_*, \end{aligned} \quad (19)$$

which makes use of Eq. (17).

Some remarks on the limitations of Eq. (19) are in order. As mentioned in the discussion leading to Eq. (1), the local plane of polarization at position $\mathbf{r} = \mathbf{p}$ must have a dependence on turbulent modulation. As a thought experiment, consider atmospheric fluctuations in the telescope entrance pupil plane in a small vicinity of the point \mathbf{p} at some time t_0 resulting in tilt only, so that $\phi_a(\mathbf{r}, t_0)|_{\mathbf{p}} = \alpha x$ [where $\mathbf{r} = (x, y)$, and $\tan^{-1}\alpha$ is the tilt angle]. Since the tilt changes the direction of propagation, an unpolarized beam would have an electric field component parallel to the z -axis, while the x -component would be correspondingly reduced. Consider that typical seeing conditions in a world-class observatory correspond deviations in the angle of propagation on the order of 1 arc second ($''$). A deviation $\alpha = 1''$ would correspond to a z -component of the electric field amplitude of about 5×10^{-6} time the x -component, but impact on the x -component itself would only be about 1 part in 10^{11} . While these numbers are not large, one must keep in mind that the telescope has optics that reduce the beam diameter while amplifying the angle. Consider an ELT that reduces the beam from a primary diameter of 50 m to a diameter of 2 cm (a factor of 2500) by the time it impinges on the DM, making $\alpha = 2500'' \approx 0.7^\circ$ on the internal pupil plane corresponding to the DM. Thus, the atmospheric fluctuations combined with the high demagnification correspond to a local beam wobble on the order of 1° in the vicinity of the conjugate point of \mathbf{p} . This deviation now imparts z -component of over 1% of the original x -component of the electric field amplitude, which itself would be reduced by nearly 1 part in 10^4 . Such considerations may not be negligible in precision optical modeling and require further investigation.

Using Eq. (19), the star's mutual coherency at the entrance pupil is:

$$\gamma_{*0}(\mathbf{r}_1, \mathbf{r}_2, t) = \gamma_{*-1}(\mathbf{r}_1, \mathbf{r}_2) \exp[j[\phi_a(\mathbf{r}_1, t) - \phi_a(\mathbf{r}_2, t)]] , \quad (20)$$

$$= J_* \exp[j[\phi_a(\mathbf{r}_1, t) - \phi_a(\mathbf{r}_2, t) + k\boldsymbol{\alpha}_* \cdot (\mathbf{r}_1 - \mathbf{r}_2)]] \quad (21)$$

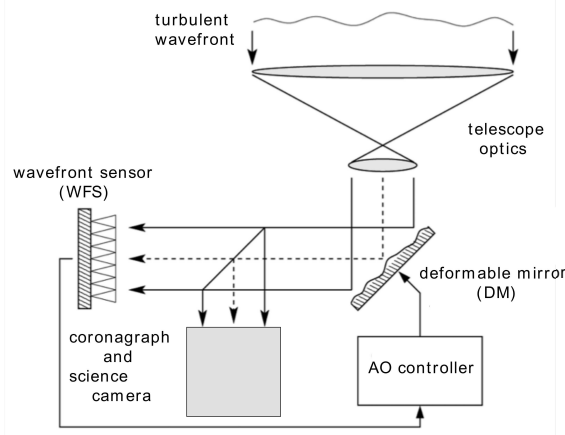


Figure 3. Schematic diagram of an optical system containing a telescope with a closed-loop AO system and a coronagraph. Modified from [37].

where Eq. (21) uses Eq. (16). From Eq. (21), the star's coherency at the entrance pupil is simply:

$$J_{*0}(\mathbf{r}, t) = J_* \exp \left[-2\Im(\phi_a(\mathbf{r}, t)) \right]. \quad (22)$$

To calculate the mutual coherency of the planetary light, one can assume that its angular size is small enough that anisoplanatism effects are negligible, so that the planetary field is modulated by the same factor $\exp[j\phi_a(\mathbf{r}, t)]$. Any small deviations from this assumption should be inconsequential, as they would result in a slight blurring of the planetary component of the image, not the stellar component. As there is no convenient factorization of the mutual coherency of the planetary light in the -1 plane above the atmosphere, as in Eq. (17), one can multiply u_{p-1} in Eq. (12) by $\exp[j\phi_a(\mathbf{r}, t)]$ and follow the same steps to arrive at a blurred version of Eq. (14) (which is simple because the atmospheric modulation factor does not depend on α), to find:

$$\gamma_{p0}(\mathbf{r}_1, \mathbf{r}_2, t) = \frac{1}{\pi} \exp j[\phi_a(\mathbf{r}_1, t) - \phi_a(\mathbf{r}_2, t)] \int d\alpha S(\alpha) \exp[jk\alpha \cdot (\mathbf{r}_1 - \mathbf{r}_2)]. \quad (23)$$

Eq. (23) shows that the Fourier transform of the polarimetric planetary image, $S(\alpha)$, is multiplied by a complex-valued atmospheric modulation function, $\exp j[\phi_a(\mathbf{r}_1, t) - \phi_a(\mathbf{r}_2, t)]$. Thus, in an imaging system, which must have Fourier transforming properties, the $S(\alpha)$ will be convolved with the Fourier transform of $\exp j[\phi_a(\mathbf{r}_1, t) - \phi_a(\mathbf{r}_2, t)]$, resulting in an image that is distorted by the atmospheric turbulence.

3.3 Through the Optical System

Consider a ground-based telescope equipped with a WFS and coronagraph, as shown schematically in Fig. 3. The telescope, coronagraph and the science camera comprise a system consisting of a number of optical surfaces, with the telescope entrance pupil corresponding to surface number 0 and the science camera detector surface having index C . Let $\Upsilon_{l+1,l}(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l)$, where \mathbf{r}_l and \mathbf{r}_{l+1} are the coordinate vectors in planes l and $l+1$, respectively, be a Jones propagator from surface l to surface $l+1$, including any aperture geometry or other structure on surface l . The $\Upsilon_{l+1,l}$ kernel explicitly excludes interaction with the $l+1$ surface. The vector $\boldsymbol{\theta}_l$ is a set of experimentally determined parameters that modify $\Upsilon_{l+1,l}$ to account for various conditions such as alignment or other possibly significant factors.

The field on the plane $l+1$ before interacting with the $l+1$ surface is given by integrating u_l against the propagation kernel matrix:

$$u_{l+1}(\mathbf{r}_{l+1}, t) = \int d\mathbf{r}_l \Upsilon_{l+1,l}(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l) u_l(\mathbf{r}_l, t). \quad (24)$$

In order to be as general as possible and allow for unknown (scalar and polarization) aberration at each optical surface, the propagation kernel can be factored into known and unknown parts as follows:

$$\Upsilon_{l+1,l}(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l) = \Upsilon_{l+1,l}^k(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l) A_l^u(\mathbf{r}_l) \quad (25)$$

where the "k" superscript stands for "known" and the "u" superscript stands for "unknown," $A_l^u(\mathbf{r}_l)$ is a 2×2 Jones matrix function of unknown (possibly complex-valued) functions representing the unknown aberration caused by surface l . If the A_l^u function were parameterized, it could be included in the definition of the $\boldsymbol{\theta}_l$ vector, as an alternative formulation. Accounting for unknown aberrations in the form of Eq. (25) provides a bit of mathematical convenience as A_l^u can be regarded as a (small) deviation from the identity, i.e.:

$$A_l^u(\mathbf{r}_l) = \mathbb{I} + \tilde{A}_l^u(\mathbf{r}_l), \quad (26)$$

where \tilde{A}_l^u is the deviation from the identity matrix (\mathbb{I}). $A_l^u(\mathbf{r}_l)$ is quite suitable for representing unknown scalar phase aberrations, such as small bumps on a mirror that have negligible effect on the polarization. Such a scalar aberration can be written as $A_l^u(\mathbf{r}_l) = \mathbb{I} \exp(j\phi_l^u(\mathbf{r}_l))$, where ϕ_l^u is the (possibly complex-valued) phase aberration on the optical surface l . Using the Taylor expansion of the exponential $(1 + j\phi_l^u(\mathbf{r}_l) - \phi_l^{u2}(\mathbf{r}_l)/2 + \dots)$, one has $\tilde{A}_l^u = \mathbb{I}(j\phi_l^u(\mathbf{r}_l) - \phi_l^{u2}(\mathbf{r}_l)/2 + \dots)$. Then, using Eq. (25), the propagation kernel including an unknown a scalar aberration in plane l can be written as:

$$\Upsilon_{l+1,l}(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l) = \Upsilon_{l+1,l}^k(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l) \mathbb{I}[1 + j\phi_l^u(\mathbf{r}_l) - \phi_l^{u2}(\mathbf{r}_l)/2 + \dots]. \quad (27)$$

The choice of the kernel matrix $\Upsilon_{l+1,l}$ may be the Jones matrix function corresponding to an optical element in plane l combined with some approximation of the Huygens-Fresnel principle to propagate the field to plane $l+1$. For example, if the 2×2 Jones matrix function for plane l is given by $A_l(\mathbf{r}_l; \boldsymbol{\theta}_l)$ (which, if desired, can be factored into known and unknown parts as $A_l^k A_l^u$), the corresponding Huygens-Fresnel kernel can be written as:

$$\Upsilon_{l,l+1}(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l) = \frac{-jk}{2\pi} A_l(\mathbf{r}_l; \boldsymbol{\theta}_l) \exp[jkr(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l)] \frac{\cos \vartheta(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l)}{r(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l)}, \quad (28)$$

where $r(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l)$ is the Euclidean distance between the point \mathbf{r}_l on surface l and \mathbf{r}_{l+1} on surface $l+1$, and $\vartheta(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l)$ is the angle between the aperture normal on surface l and the line connecting these two points.

One can propagate the field from surface l to surface $l+2$ by using the kernel $\Upsilon_{l+2,l}$, which is defined as the contraction of the two kernels $\Upsilon_{l+1,l}$ and $\Upsilon_{l+2,l+1}$:

$$\Upsilon_{l+2,l}(\mathbf{r}_{l+2}, \mathbf{r}_l; \boldsymbol{\theta}_{l+1}, \boldsymbol{\theta}_l) \equiv \int d\mathbf{r}_{l+1} \Upsilon_{l+2,l+1}(\mathbf{r}_{l+2}, \mathbf{r}_{l+1}; \boldsymbol{\theta}_{l+1}) \Upsilon_{l+1,l}(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l). \quad (29)$$

In operator notation Eq. (29) is written as:

$$\Upsilon_{l+2,l}(\mathbf{r}_{l+2}, \mathbf{r}_l; \boldsymbol{\theta}_{l+1}, \boldsymbol{\theta}_l) \equiv \Upsilon_{l+2,l+1}(\mathbf{r}_{l+2}, \mathbf{r}_{l+1}; \boldsymbol{\theta}_{l+1}) \Upsilon_{l+1,l}(\mathbf{r}_{l+1}, \mathbf{r}_l; \boldsymbol{\theta}_l), \quad (30)$$

which implies integration over the coordinate that is the second argument in the first propagator and the first argument in the second propagator.

Similarly, the field from the star at the science camera, located in plane C can be determined from the propagation kernel $\Upsilon_{C,0}$, which is a contraction of the propagation kernels from plane 1 through plane $C-1$:

$$\Upsilon(\mathbf{r}_C, \mathbf{r}_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:C-1}) \equiv \int d\mathbf{r}_{C-1} \dots \int d\mathbf{r}_1 \Upsilon_{C,C-1}(\mathbf{r}_C, \mathbf{r}_{C-1}; \boldsymbol{\theta}_{C-1}) \dots \Upsilon_{1,0}(\mathbf{r}_1, \mathbf{r}_0; \boldsymbol{\theta}_0), \quad (31)$$

where $\mathbf{m}(t)$ is a vector of DM commands, $\boldsymbol{\theta}_{0:C} \equiv [\boldsymbol{\theta}_0^T, \dots, \boldsymbol{\theta}_C^T]^T$ is a column vector of instrument parameters that collects the $\{\boldsymbol{\theta}_l\}$. Note that in Eq. (31) hides the dependence on $\mathbf{m}(t)$ insides the ellipsis ("..."), which contains a propagation kernel corresponding to the DM that depends on $\mathbf{m}(t)$. In operator notation, Eq. (31) is written as:

$$\Upsilon(\mathbf{r}_C, \mathbf{r}_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:C-1}) = \prod_{l=0}^{C-1} \Upsilon_{l+1,l}(\mathbf{r}_{l+1}, \mathbf{r}_l; \mathbf{m}(t); \boldsymbol{\theta}_l), \quad (32)$$

in which integrations over the intermediate coordinate vectors are implicit.

Using Eq. (19), the field from the star at the science camera is given by integral:

$$u_C(\boldsymbol{\rho}, t) = \sqrt{I_\star} \int d\mathbf{r}_0 \Upsilon_{C,0}(\boldsymbol{\rho}, \mathbf{r}_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:C-1}) \check{u}_\star \exp j[k\boldsymbol{\alpha}_\star \cdot \mathbf{r}_0 + \phi_a(\mathbf{r}_0, t)], \quad (33)$$

where the position in science camera plane is given by $\boldsymbol{\rho} \equiv \mathbf{r}_C$. The polarimetric image of the star can be determined with the help of Eqs. (11) and (21):

$$J_{\star C}(\boldsymbol{\rho}, t) = \int d\mathbf{r}_0 \int d\mathbf{r}'_0 \Upsilon_{C,0}(\boldsymbol{\rho}, \mathbf{r}_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:C}) \otimes \Upsilon_{C,0}^*(\boldsymbol{\rho}, \mathbf{r}'_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:C}) \times \\ J_\star \exp j[\phi_a(\mathbf{r}_0, t) - \phi_a(\mathbf{r}'_0, t) + k\boldsymbol{\alpha}_\star \cdot (\mathbf{r}_0 - \mathbf{r}'_0)]. \quad (34)$$

Using Eq. (23), the polarimetric image of the planetary system (excluding the star) is given by:

$$J_{pC}(\boldsymbol{\rho}, t) = \int d\mathbf{r}_0 \int d\mathbf{r}'_0 \exp j[\phi_a(\mathbf{r}_0, t) - \phi_a(\mathbf{r}'_0, t)] \Upsilon_{C,0}(\boldsymbol{\rho}, \mathbf{r}_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:C}) \otimes \Upsilon_{C,0}^*(\boldsymbol{\rho}, \mathbf{r}'_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:C}) \times \\ \int_p d\boldsymbol{\alpha} S(\boldsymbol{\alpha}) \exp [k\boldsymbol{\alpha} \cdot (\mathbf{r}_0 - \mathbf{r}'_0)]. \quad (35)$$

Note that detailed propagation kernels in Eq. (35) are likely to be unnecessarily precise, as much more simple approximations to calculate the planetary image should be adequate. This will be discussed in future papers in this series. Adding Eqs. (34) and (35), the polarimetric image on the science camera detector is the sum of the stellar and planetary contributions:

$$J_C(\boldsymbol{\rho}, t) = J_{\star C}(\boldsymbol{\rho}, t) + J_{pC}(\boldsymbol{\rho}, t). \quad (36)$$

Using the operator notation as in Eq. (50), $\Upsilon_{C,0}(\boldsymbol{\rho}, \mathbf{r}_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:C}) \otimes \Upsilon_{C,0}^*(\boldsymbol{\rho}, \mathbf{r}'_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:C})$ also represents the operator that propagates the mutual coherency through the optical system. Using Eq. (32) and the identity (for suitably dimensioned matrices) $AC \otimes BD = (A \otimes B)(C \otimes D)$, this operator can be written as:

$$\Upsilon_{C,0}(\boldsymbol{\rho}, \mathbf{r}_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:C}) \otimes \Upsilon_{C,0}^*(\boldsymbol{\rho}, \mathbf{r}'_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:C}) \\ = \prod_{l=0}^{C-1} [\Upsilon_{l+1,l}(\mathbf{r}_{l+1}, \mathbf{r}_l; \mathbf{m}(t); \boldsymbol{\theta}_l) \otimes \Upsilon_{l+1,l}^*(\mathbf{r}'_{l+1}, \mathbf{r}'_l; \mathbf{m}(t); \boldsymbol{\theta}_l)], \quad (37)$$

in which $\mathbf{r}'_C = \mathbf{r}_C = \boldsymbol{\rho}$, and the $\mathbf{m}(t)$ (the DM command vector) argument only affects the operators in which the second coordinate corresponds to a DM.

Referring to Fig. 3, let the kernel that propagates the field from the telescope entrance pupil to the WFS entrance pupil be denoted as $\Upsilon_{w,0}(\mathbf{r}_w, \mathbf{r}_0; \boldsymbol{\theta}_{0:w-1})$, where w is the index assigned to the plane corresponding to the WFS entrance pupil. Neglecting the planetary contribution, the field in the WFS entrance pupil is given by:

$$u_w(\mathbf{r}_w, t) = \sqrt{I_\star} \int d\mathbf{r}_0 \Upsilon_{w,0}(\mathbf{r}_w, \mathbf{r}_0; \mathbf{m}(t); \boldsymbol{\theta}_{0:w-1}) \check{u}_\star \exp j[k\boldsymbol{\alpha}_\star \cdot \mathbf{r}_0 + \phi_a(\mathbf{r}_0, t)], \quad (38)$$

In Part 2 of this series, the SC field will be expressed in terms of u_w and non-common path aberrations, which are unique to the separate optical path experienced by the field impinging on the WFS.

4. METHODS FOR TREATMENT OF ABERRATIONS

The purpose of this section is to outline the physical principles behind various aberration correction schemes in a rigorous fashion. The level of rigor will allow the reader to understand the strengths and weaknesses of the various approaches. This section does not discuss issues concerning noise, measurement errors, statistical or computational methodologies. These critical concerns are left to later papers in the series.

4.1 Explicit Determination of Unknown Aberrations

The first step in solving for the unknown aberrations is to calculate how they manifest themselves in the polarimetric image. Using Eqs. (25) and (26), and dropping terms that are 2nd order or higher in $\{\tilde{A}_l\}$, Eq. (32) becomes:

$$\Upsilon_{C,0}(\mathbf{r}_C, \mathbf{r}_0; \mathbf{m}(t)) = \prod_{l=0}^{C-1} \Upsilon_{l+1,l}^k(\mathbf{r}_{l+1}, \mathbf{r}_l; \mathbf{m}(t)) + \sum_{k=1}^{C-1} \left(\prod_{l=k+1}^{C-1} \Upsilon_{l+1,l}^k(\mathbf{r}_{l+1}, \mathbf{r}_l; \mathbf{m}(t)) \right) \Upsilon_{k+1,k}^k(\mathbf{r}_{k+1}, \mathbf{r}_k; \mathbf{m}(t)) \tilde{A}_k^u(\mathbf{r}_k) \left(\prod_{l=0}^{k-1} \Upsilon_{l+1,l}^k(\mathbf{r}_{l+1}, \mathbf{r}_l; \mathbf{m}(t)) \right) \quad (39)$$

where, for the sake of compactness, the $\{\theta_l\}$ arguments, have been dropped, and in which the roman superscript k , which stands for the word “known,” should not be confused with the index k (let alone the wavenumber $k = 2\pi/\lambda$). In Eq. (39), the first term is simply the known part of the propagation operator, while the second term provides a 1st order accounting of the surface aberrations $\{\tilde{A}_k(\mathbf{r}_k)\}$ by summing over the effect of each surface. In the second term, the rightmost factor propagates the field from the entrance pupil to the k th surface, where the unknown aberration is located, the Jones matrix function $\tilde{A}_k^u(\mathbf{r}_k)$ applies the aberration, $\Upsilon_{k+1,k}$ propagates the field to the $k+1$ surface, and the leftmost factor propagates the field the rest of the way to the science camera detector surface. Eq. (39) assumes that there are no unknown aberrations in at surfaces 0 and C . The first-order image plane manifestation of the unknown aberrations can be calculated from Eq. (34), using the expansion in Eq. (39), dropping terms that are 2nd order in the $\{\tilde{A}_k\}$ functions. In operator notation, the resulting polarimetric image of the star is:

$$J_{\star C}(\boldsymbol{\rho}, t) \approx \left\{ \Upsilon_{C,0}^k(\boldsymbol{\rho}, \mathbf{r}_0; \mathbf{m}(t)) \otimes \Upsilon_{C,0}^{k*}(\boldsymbol{\rho}, \mathbf{r}'_0; \mathbf{m}(t)) + \Upsilon_{C,0}^k(\boldsymbol{\rho}, \mathbf{r}_0; \mathbf{m}(t)) \otimes \left[\sum_{k=1}^{C-1} \left(\prod_{l=k+1}^{C-1} \Upsilon_{l+1,l}^{k*}(\mathbf{r}'_{l+1}, \mathbf{r}'_l; \mathbf{m}(t)) \right) \Upsilon_{k+1,k}^{k*}(\mathbf{r}'_{k+1}, \mathbf{r}'_k; \mathbf{m}(t)) \tilde{A}_k^{u*}(\mathbf{r}'_k) \left(\prod_{l=0}^{k-1} \Upsilon_{l+1,l}^{k*}(\mathbf{r}'_{l+1}, \mathbf{r}'_l; \mathbf{m}(t)) \right) \right] + \left[\sum_{k=1}^{C-1} \left(\prod_{l=k+1}^{C-1} \Upsilon_{l+1,l}^k(\mathbf{r}_{l+1}, \mathbf{r}_l; \mathbf{m}(t)) \right) \Upsilon_{k+1,k}^k(\mathbf{r}_{k+1}, \mathbf{r}_k; \mathbf{m}(t)) \tilde{A}_k^u(\mathbf{r}_k) \left(\prod_{l=0}^{k-1} \Upsilon_{l+1,l}^k(\mathbf{r}_{l+1}, \mathbf{r}_l; \mathbf{m}(t)) \right) \right] \otimes \Upsilon_{C,0}^{k*}(\boldsymbol{\rho}, \mathbf{r}'_0; \mathbf{m}(t)) \right\} \times J_{\star} \exp j[\phi_a(\mathbf{r}_0, t) - \phi_a(\mathbf{r}'_0, t) + k\boldsymbol{\alpha}_{\star} \cdot (\mathbf{r}_0 - \mathbf{r}'_0)] , \quad (40)$$

where $\Upsilon_{C,0}^k$ is the first term in Eq. (39), corresponding to the known propagator, and $\mathbf{r}_C = \mathbf{r}'_C = \boldsymbol{\rho}$. Similarly to Eq. (37), the operators constituting $\Upsilon_{C,0}^k$ may be placed inside the serial products over the indices l . Eq. (40) is linear in unknown aberration functions $\{\tilde{A}_k(\mathbf{r}_k)\}$ and is the fundamental equation for statistical inference of their values, under the assumption that the aberrant field is small enough that the 2nd order interactions are negligible. This assumption can be improved once the largest unknown aberrations are determined by including them in the known propagation operators. It is also important to note that Eq. (40) makes no assumptions about the DM commands, which may include deliberate offsets associated with speckle nulling or electric field conjugation schemes, as discussed in Sec. 44.3.

In Part 2 of this series, it will be shown that the SC image can be written in terms of u_w , the field measured by the WFS. The resulting expression eliminates the need to consider aberrations upstream of the beam splitter, which separates the optical paths of the SC and WFS, and the result is a versions of Eqs. (39) and (40) with many fewer unknown functions.

Of course, the image in the science camera also includes the planetary image, as per Eqs. (35) and (36). The known propagation operator $\Upsilon_{C,0}^k$ should be more than adequate for calculating the planetary image. Thus, when one combines Eq. (40) with a version Eq. (35) that utilizes $\Upsilon_{C,0}^k$, the resulting expression for the polarimetric image in the science camera is linear in both the unknown aberration functions and the planetary image $S(\boldsymbol{\alpha})$. Then, the exoplanet imaging problem in the fairly general form developed here can be stated as: Given data

streams corresponding to measurements of the residual phase $\phi_r(\mathbf{r}, t)$ in Eq. (38) and $J_C(\boldsymbol{\rho}, t)$ in Eq. (36), estimate $S(\boldsymbol{\alpha})$, $\boldsymbol{\theta}_{0:C-1}$, and $\{\tilde{A}_l^u(\mathbf{r}_l)\}$. As the image J_C is strongly modulated by the atmospheric fluctuations ϕ_a , the use of the WFS data stream is critical to success in this endeavor.^{11,13} While estimating so many unknown functions is certainly a daunting challenge, it should not be dismissed so readily, as large millisecond-cadence data sets, modern hardware, and computational/statistical methods (e.g., sparse representations) could make significant strides in this direction a realistic long-term goal. In the short term, simplifications must be considered, which will be treated in future papers in this series.

4.1.1 The Concept of Equivalent Aberrations

Determining unknown aberration functions on many optical surfaces simultaneously is an ambitious goal, and finding methods to reduce the number of unknown functions that need to be determined is worth some effort. Consider a hypothetical optical system with unknown aberrations in only two planes, M and N , where $N > M > D$, meaning that plane N is downstream of plane M , which is downstream of the DM plane D . Assume that the M and N planes have small scalar aberrations, so that, as per Eq. (26), $\tilde{A}_M^u = \mathbb{I}j\phi_M^u(\mathbf{r}_M)$ and $\tilde{A}_N^u = \mathbb{I}j\phi_N^u(\mathbf{r}_N)$. Under what circumstances can one account for the aberration in plane M by replacing it with some equivalent aberration in plane N ? Using Eq. (27), the aberrant part of the field incident on plane N is:

$$u_N^u(\mathbf{r}_N, t) = j \int d\mathbf{r}_M \Upsilon_{N,M}^k(\mathbf{r}_N, \mathbf{r}_M) \phi_M^u(\mathbf{r}_M) u_M(\mathbf{r}_M, t) \quad (41)$$

assuming that there are no active optical elements (i.e., DMs) between M and N . Replacing $\phi_M^u(\mathbf{r}_M)$ with some functionally equivalent aberration in the N plane equates making the following assertion: There exists some complex-valued and time-independent function $\psi(\mathbf{r}_N)$ such that, for all (reasonable) values of $u_M(\mathbf{r}_M, t)$ and $\phi_M^u(\mathbf{r}_M)$, the following relation holds:

$$\psi(\mathbf{r}_N) \int d\mathbf{r}_M \Upsilon_{N,M}^k(\mathbf{r}_N, \mathbf{r}_M) u_M(\mathbf{r}_M, t) = \int d\mathbf{r}_M \Upsilon_{N,M}^k(\mathbf{r}_N, \mathbf{r}_M) \phi_M^u(\mathbf{r}_M) u_M(\mathbf{r}_M, t). \quad (42)$$

If this relation holds true, the $k = M$ term in Eq. (39) can be omitted if the $k = N$ term is modified according to $\phi_N^u(\mathbf{r}_N) \rightarrow [\phi_N^u(\mathbf{r}_N) + \psi(\mathbf{r}_N)]$. From Eq. (42), $\psi(\mathbf{r}_N)$ would be given by the ratio:

$$\psi(\mathbf{r}_N) = \frac{\int d\mathbf{r}_M \Upsilon_{N,M}^k(\mathbf{r}_N, \mathbf{r}_M) \phi_M^u(\mathbf{r}_M) u_M(\mathbf{r}_M, t)}{\int d\mathbf{r}_M \Upsilon_{N,M}^k(\mathbf{r}_N, \mathbf{r}_M) u_M(\mathbf{r}_M, t)}, \quad (43)$$

possible zeros in the denominator notwithstanding. In order for the time-independent function $\psi(\mathbf{r}_N)$ to exist, one of several conditions must be met:

- There are no unknown aberrations, so $\phi_M^u = 0$.
- The incident field is not time-dependent so $u_M(\mathbf{r}_M, t) = u_M(\mathbf{r}_M)$. If the M plane is after the DM, condition is approached as the quality of the AO correction improves, i.e., when a high Strehl ratio is achieved.
- The propagation operator $\Upsilon_{N,M}^k$ behaves much like an identity operator, so that for some arbitrary field $u_M(\mathbf{r}_M)$, (in operator notation) $u_N(\mathbf{r}_N) = \Upsilon_{N,M}^k(\mathbf{r}_N, \mathbf{r}_M) u_M(\mathbf{r}_M) = u_M(\beta \mathbf{r}_N)$ for some (real-valued) magnification factor β , in which case $\psi(\mathbf{r}_N) = \phi_M^u(\beta \mathbf{r}_M)$. Under geometrical optics approximations, this condition is achieved when the M and N planes are conjugate.

There may be other, most likely contrived, functional forms of $\Upsilon_{N,M}^k$ and ϕ^u in which $\psi(\mathbf{r}_N)$ exists. The general formula for propagating the field from one surface to the next involves the Huygen-Fresnel kernel, shown in Eq. (28), which does not behave like the identity operator unless intermediate optical surfaces create 2^n (where n is a natural number) Fourier transforms, or the beam is columnated and propagation effects can be neglected. Therefore, it seems that the treating aberrations on multiple optical surfaces equivalently to aberrations on a single optical surface, is an approximation that needs to be evaluated on a case-by-case basis.

4.2 Black Box Models

The developments thus far allow the reader to understand the speckle cancellation method of [38] and the background estimation method of [12]. As per Eq. (39), the end-to-end propagator can be written as the sum of known and unknown propagators: $\Upsilon_{C,0} = \Upsilon_{C,0}^k + \Upsilon_{C,0}^u$, where the unknown propagator is the second term in Eq. (39), plus terms that are nonlinear in the $\{\tilde{A}_k\}$, if desired. Here, this is called the "black box model" since the unknown aberrations are only taken into account via the resulting image-plane field u^u . Black box models have the considerable advantage of not requiring a sophisticated model of the optical system and a parameterization of the unknown aberrations. Assume that the WFS data stream leads to an estimate of the atmospheric modulation $\exp[j\hat{\phi}_a(\mathbf{r}_0, t)]$. Then, the field in the science camera is given by:

$$u_C(\boldsymbol{\rho}, t) \approx u^k(\boldsymbol{\rho}, t) + u^u(\boldsymbol{\rho}, t), \quad (44)$$

in which the known and unknown parts of the field are given by [see Eq. (33)]:

$$u^k(\boldsymbol{\rho}, t) = \sqrt{I_\star} \int d\mathbf{r}_0 \Upsilon_{C,0}^k(\mathbf{r}_C, \mathbf{r}_0; \mathbf{m}(t)) \check{u}_\star \exp[j\hat{\phi}_a(\mathbf{r}_0, t)] \quad (45)$$

$$u^u(\boldsymbol{\rho}, t) = \sqrt{I_\star} \int d\mathbf{r}_0 \Upsilon_{C,0}^u(\mathbf{r}_C, \mathbf{r}_0; \mathbf{m}(t)) \check{u}_\star \exp[j\hat{\phi}_a(\mathbf{r}_0, t)] \quad (46)$$

in which $\boldsymbol{\alpha}_\star = 0$ has been assumed, and the $\boldsymbol{\theta}_{0:C}$ argument has been dropped for simplicity. The method of [12] requires the unknown field u^u ("the subject beam") to be independent of the atmospheric modulation (and therefore time), i.e., $u^u(\boldsymbol{\rho}) \approx \sqrt{I_\star} \int d\mathbf{r}_0 \Upsilon_{C,0}^u(\mathbf{r}_C, \mathbf{r}_0) \check{u}_\star$, which is not consistent with Eq. (46). Then, ignoring the planetary contribution and polarization effects, the image in the science camera is given by

$$\begin{aligned} I_C(\boldsymbol{\rho}, t) & (=) \quad u_C(\boldsymbol{\rho}, t) u_C^*(\boldsymbol{\rho}, t) \\ & (=) \quad u^k(\boldsymbol{\rho}, t) u^{k*}(\boldsymbol{\rho}, t) + 2\Re[u^u(\boldsymbol{\rho}) u^{k*}(\boldsymbol{\rho}, t)] + u^u(\boldsymbol{\rho}) u^{u*}(\boldsymbol{\rho}), \end{aligned} \quad (47)$$

where I_C is the intensity measured in the science camera, u^u and u^k are taken to be scalar fields, \Re indicates that the real part is to be taken, and the $=$ is placed inside parentheses because the equation is based on the incorrect assumption discussed above. Given a time-series of measurements of $J_C(\boldsymbol{\rho}, t)$ and estimates $\hat{\phi}_a(\mathbf{r}, t)$, Eq. (47) allows determination $u^u(\boldsymbol{\rho})$ and, therefore, the intensity in the image plane arising due to unknown aberrations, i.e., $2\Re[u^u(\boldsymbol{\rho}) u^{k*}(\boldsymbol{\rho}, t)] + u^u(\boldsymbol{\rho}) u^{u*}(\boldsymbol{\rho})$. With $u^k(\boldsymbol{\rho}, t)$ and $u^u(\boldsymbol{\rho})$ in hand, the background to be subtracted from the target image is simply the time-integral of $\|u^k(\boldsymbol{\rho}, t) + u^u(\boldsymbol{\rho})\|^2$.

Of course, the major criticism of this approach is the assumption that the field arising from the unknown aberrations is independent of the turbulent modulation, which, clearly, it is not, according to Eq. (46). Indeed, the real part ϕ_a may well vary over multiples of 2π over the pupil. However, this assumption is improved by the fact that, for planes in the optical system succeeding the DM, the uncorrected portion of ϕ_a , the residual phase ϕ_r , has a significantly smaller amplitude. It is likely that the unknown aberrations are most significant after the DM, because, in a closed-loop system, the AO system corrects most of the aberrations up to the DM plane.

There may be a possibility to improve the method of [12] for closed-loop AO systems by assuming that the unknown portion of the field in the image plane is a function only of the residual phase $\phi_r(\mathbf{r}, t)$, so $u_u(\boldsymbol{\rho}) \rightarrow u^u(\boldsymbol{\rho}; \phi_r(\mathbf{r}, t))$. Then for small amplitudes of $\phi_r(\mathbf{r}, t)$, one has

$$u^u(\boldsymbol{\rho}; \phi_r(\mathbf{r}, t)) \approx u^u(\boldsymbol{\rho}; 0) + \left. \frac{\partial u^u(\boldsymbol{\rho}; \phi_r(\mathbf{r}, t))}{\partial \phi_r} \right|_{\phi_r=0} \hat{\phi}_r(\mathbf{r}, t), \quad (48)$$

which makes use of the functional derivative and WFS estimate of the residual phase $\hat{\phi}_r$. Then one could replace $u^u(\boldsymbol{\rho})$ in Eq. (47) with $u^u(\boldsymbol{\rho}; \phi_r(\mathbf{r}, t))$ in Eq. (48), and use measurements to solve for both $u^u(\boldsymbol{\rho}; 0)$ and a representation of the functional derivative $\left. \frac{\partial u^u(\boldsymbol{\rho}; \phi_r(\mathbf{r}, t))}{\partial \phi_r} \right|_{\phi_r=0}$. The difficulty of this scheme increases dramatically if higher order functional derivatives are needed to approve the approximation in Eq. (48).

The speckle cancellation method of [38] does not make use of the WFS data stream or millisecond exposures in the science camera. The goal of the speckle cancellation procedure is to use the DM to cause starlight to

destructively interfere with an undesired speckle at position $\boldsymbol{\rho}$. This procedure is repeated to remove multiple speckles at various locations. In this method one seeks to cancel out a speckle in the image plane at the location $\boldsymbol{\rho}$, caused by an aberration leading to speckle field $u^u(\boldsymbol{\rho})$, making the same assumption as [12], i.e., ignoring its dependence on the atmospheric turbulence, in violation of Eq. (46). In this method, relatively long exposures are used, so the observations are effectively time-averages over the atmospheric turbulence. The principle of the method is to use a several known DM offsets \mathbf{m}_k , in which the index $k = (0, 1, 3)$ corresponds to the phases $\{\varphi_k\} = (0, \pi/2, \pi, 3\pi/2)$ of a sinusoidal DM pattern corresponding to the spatial frequency $\boldsymbol{\rho}$. Then, at a given time t the DM position corresponds to $\mathbf{m}(t) + \mathbf{m}_k$, in which $\mathbf{m}(t)$ is the set of commands determined by the AO control loop. Then the known part of the field is $u^k(\boldsymbol{\rho}, t; \varphi_k)$, which can be calculated from Eq. (45) using $\mathbf{m}(t) + \mathbf{m}_k$ as an argument inside $\Upsilon_{C,0}^k$ instead of $\mathbf{m}(t)$. As long exposures are used, one measures $\overline{I_C}(\boldsymbol{\rho}, t; \varphi_k)$, where the overline indicates temporal averaging over the atmospheric turbulence, to determine the amplitude and phase of $u^u(\boldsymbol{\rho})$. Once the amplitude and phase of $u^u(\boldsymbol{\rho})$ have been determined, the DM is commanded to implement a canceling sinusoidal aberration $\delta\mathbf{m}$ that gives rise to a field $u^c(\boldsymbol{\rho})$ that is equal in amplitude, but opposite in phase to, $u^u(\boldsymbol{\rho})$. In reality neither u^u nor u^c will be independent of the atmospheric modulation, but some cancellation can be achieved, and [39] demonstrates considerable success on-sky with this technique.

4.3 Dark Hole Generation & Electric Field Conjugation

Considerable effort has been devoted to using the DM (or multiple DMs) to minimize (or otherwise control) the starlight in some specified region of the image plane. The history of these efforts is reviewed in [22] and [38]. The idea behind so-called "dark hole" methods is to minimize the intensity integrated over a control region of interest,²¹ which requires first determining the speckle field using prescribed DM offsets, much as described above in Sec. 4.4.2 in connection with the speckle cancellation method of [38]. So-called "electric field conjugation" does not seek to minimize the intensity in the control region. Rather, its objective is to make the field what it would be in the absence of unknown aberration, or, in other words, to attain some ideal diffraction pattern. Modern implementations set the DM offset command vector $\delta\mathbf{m}$ to a value that minimizes the required DM stroke while the desired image plane electric field is used as a constraint on the optimization.^{22, 40}

Dark hole and field conjugation methods assume that the unwanted speckle in the image plane is caused by a complex-valued unknown aberration in some plane of the optical system, most often a pupil plane corresponding to the location of a DM. Let us denote this plane with the index N . If the unknown aberrations in the optical system are truly confined only to the N plane, then the field in the science camera is determined by the propagation kernel in Eq. (39), keeping only the $k = N$ term in the summation. Thus, one fundamental limitation with these methods is that they ignore the $k \neq N$ terms, and implicitly assume that all of the unaccounted aberration can be treated by an equivalent pupil plane aberration, as discussed in Sec. 4.4.1.1. In observations from space, the assumption of equivalent aberrations may be justified, but the ratio in Eq. (43) will have some (relatively slow) time-dependence due to telescope dynamics, requiring the DM solution to the field conjugation problem to be updated regularly.⁴⁰ If these updates are accurate and on a time-scale that is short enough, the approximation of equivalent aberrations should not prove to be a problem. In ground-based astronomy, this approximation of equivalent aberrations needs to be evaluated with much greater care and will depend not only on propagators between the planes but also on the assumed quality of the AO correction.

One potential improvement would be to generalize dark hole and field conjugation methods to treat aberrations in multiple planes using multiple DMs, but this has not been explored, to the best of this author's knowledge. Using a complicated series of DM offsets to determine the speckle fields arising from aberrations in multiple planes may well introduce non-trivial complications, but perhaps such an approach could be unified with the methods described in this series of papers.

5. CONCLUSIONS

Direct imaging of exo-planetary systems from ground-based telescopes is fraught with technical challenges, but continuing advances in optical technology such as ultra-low noise, rapid-readout optical and infrared detectors are opening new avenues of opportunity. Taking the greatest advantage of these technologies will require harnessing simultaneous wavefront sensor and science camera data streams at millisecond cadences to differentiate the planetary signal from the aberrant starlight. The work presented here is Part I of a statistical framework that

should prove useful for self-consistently determining the unknown aberrations and their effects on the star's image in a polarizing optical system, while simultaneously solving for the (polarimetric) planetary image. The formalism presented here provides a new perspective on other methods that utilize focal plane wavefront sensing, and shows that their simplifying assumptions may limit their effectiveness in ground-based astronomy. In particular, current dark hole and electric field conjugation techniques assume that the speckle field arises due to aberrations in some specific (pupil) plane of the optical system, tacitly presuming that all of the aberrations on the various optical surfaces can be represented as a complex-valued aberration in that plane. Sec. 44.1.1 shows that this assumption of equivalent aberrations cannot be true in a strict mathematical sense, limiting the effectiveness of these methods, but the accuracy and consequences of this assumption needs to be evaluated on case-by-case basis. In principle, reliance on the assumption of equivalent aberrations could be lessened by generalizing field conjugation and dark hole methods to correct for aberrations in multiple planes. Future techniques that use DMs to correct for multi-plane aberrations would likely benefit from the developments in this series of papers for utilizing millisecond telemetry to determine the aberrations in multiple planes. Despite the limitations of the current methods that utilize focal plane sensing, they be worth implementing to provide some correction of the speckle field, but tapping the ultimate potential of the optical system will likely require the more advanced techniques outlined herein and in future papers in this series.

An exoplanet imaging program that implements simultaneous determination of the image of the planetary system and aberrations in multiple planes, utilizing millisecond telemetry from one or more WFSs and the science camera (and possibly including active aberration correction), is the most complete solution to the ultra-high contrast imaging problem. Importantly, this solution paradigm can easily include constraints, such as those arising from the diurnal rotation (used by ADI), multi-wavelength imaging (using by SDI), and polarization of the light scattered by the planetary material (used in polarization differential imaging), leaving little reason to believe it cannot lead to a substantial improvement over current methods. The chances for success of this new paradigm are greatly increased by the new generation of ultra-low noise, rapid-readout detectors. However, practical implementation of such a program will require progress on many fronts, such as sophisticated optical modeling based on various approximations, evaluating the accuracy of those approximations, modeling the error statistics of wavefronts measured by WFSs, and development of specialized statistical signal processing algorithms. These issues will be the subject of future papers in this series.

Acknowledgments

The author thanks Olivier Guyon, Wes Traub and Jim Breckinridge for encouragement and enlightening discussions.

Appendix: Formalism and Conventions

Henceforth, this paper will adhere to the following conventions in order to minimize notation:

- All time-variation made explicit in the notation corresponds to timescales of τ_G or longer. Variations on timescales similar to τ_c or smaller are implicit and dropped from the notation.
- When a function is independent of the spatial coordinate \mathbf{r} or the time t (corresponding to timescales of τ_G or longer), the argument is dropped. The number of arguments needed will be clear from the context.
- Averaging over many periods of τ_c , as was done in Eqs. (7) and (9) will be implicit whenever Jones vectors u are multiplied together. Therefore, the $\langle \rangle_{\tau_c}$ notation is dropped.

With these conventions, the field due to an astronomical source above the atmosphere in Eq. (6) is written as $u(\mathbf{r}) = \sqrt{I} \exp[j\phi(\mathbf{r})]\check{u}(\mathbf{r})$, but its coherency is independent of position and is written as $J = u(\mathbf{r}) \otimes u(\mathbf{r})^* = I\check{u}(\mathbf{r}) \otimes \check{u}^*(\mathbf{r})$. Its mutual coherency is written as $\gamma(\mathbf{r}_1, \mathbf{r}_2) = u(\mathbf{r}_1) \otimes u^*(\mathbf{r}_2) = I \exp[j(\phi(\mathbf{r}_1) - \phi(\mathbf{r}_2))]\check{u}(\mathbf{r}_1) \otimes \check{u}^*(\mathbf{r}_2)$. Now, assume the atmosphere applies a modulation to this field in the form of a (complex-valued) phase factor $\phi_a(\mathbf{r}, t)$, which varies on the τ_G timescale (see Sec. 33.2). Then, the field observed on the ground is given by $u(\mathbf{r}, t) = \sqrt{I} \exp[j(\phi(\mathbf{r}) + \phi_a(\mathbf{r}, t))]\check{u}(\mathbf{r})$, and the corresponding coherency is $J(\mathbf{r}, t) = I \exp[-2\Im(\phi_a(\mathbf{r}, t))]\check{u}(\mathbf{r}) \otimes$

$\check{u}^*(\mathbf{r})$. Similarly, the mutual coherency on the ground is expressed as $\gamma(\mathbf{r}_1, \mathbf{r}_2, t) = I \exp j[\phi(\mathbf{r}_1) - \phi(\mathbf{r}_2) + \phi_a(\mathbf{r}_1, t) - \phi_a^*(\mathbf{r}_2, t)]\check{u}(\mathbf{r}_1) \otimes \check{u}^*(\mathbf{r}_2)$.

In some of the developments in this series explicit integral notation will prove cumbersome, so operator notation will be preferred. Overloading the Υ symbol so that it serves as both the propagation kernel as in Eq. (10) and the operator that does the propagation should not introduce confusion, as the kernel needs to be accompanied by an integral sign to propagate the field, while in the operator notation the integration is implicit.[†] In operator notation, Eq. (10) is written as:

$$u_1(\mathbf{r}_1, t) = \Upsilon(\mathbf{r}_1, \mathbf{r}_0)u_0(\mathbf{r}_0, t), \quad (49)$$

and Eq. (11) becomes

$$J_1(\mathbf{r}_1, t) = \Upsilon(\mathbf{r}_1, \mathbf{r}_0) \otimes \Upsilon^*(\mathbf{r}_1, \mathbf{r}'_0)\gamma_0(\mathbf{r}_0, \mathbf{r}'_0, t). \quad (50)$$

REFERENCES

1. O. Guyon, “Direct imaging of habitable planets with ELTs,” in “Second International Conference on Adaptive Optics for Extremely Large Telescopes. id.11,” (2011).
2. W. A. Traub and B. R. Oppenheimer, Direct Imaging of Exoplanets (2010), pp. 111–156.
3. A. Boccaletti, P. Riaud, P. Baudoz, J. Baudrand, D. Rouan, D. Gratadour, F. Lacombe, and A.-M. Lagrange, “The Four-Quadrant Phase Mask Coronagraph. IV. First Light at the Very Large Telescope,” *PASP***116**, 1061–1071 (2004).
4. P. Martinez, M. Kasper, A. Costille, J. F. Sauvage, K. Dohlen, P. Puget, and J. L. Beuzit, “Speckle temporal stability in XAO coronagraphic images. II. Refine model for quasi-static speckle temporal evolution for VLT/SPHERE,” *A&A***554**, A41 (2013).
5. C. Marois, B. Macintosh, and J.-P. Véran, “Exoplanet imaging with LOCI processing: photometry and astrometry with the new SOSIE pipeline,” in “Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series,” , vol. 7736 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series (2010), vol. 7736 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series.
6. J. Rameau, G. Chauvin, A.-M. Lagrange, A.-L. Maire, A. Boccaletti, and M. Bonnefoy, “Detection limits with spectral differential imaging data,” *A&A***581**, A80 (2015).
7. C. C. Stark, A. Roberge, A. Mandell, M. Clampin, S. D. Domagal-Goldman, M. W. McElwain, and K. R. Stapelfeldt, “Lower Limits on Aperture Size for an ExoEarth Detecting Coronagraphic Mission,” *ApJ***808**, 149 (2015).
8. R. A. Brown, “Science Parametrics for Missions to Search for Earth-like Exoplanets by Direct Imaging,” *ApJ***799**, 87 (2015).
9. D. Mawet, J. Milli, Z. Wahhaj, D. Pelat, O. Absil, C. Delacroix, A. Boccaletti, M. Kasper, M. Kenworthy, C. Marois, B. Mennesson, and L. Pueyo, “Fundamental Limitations of High Contrast Imaging Set by Small Sample Statistics,” *ApJ***792**, 97 (2014).
10. J. B. Breckinridge, W. S. T. Lam, and R. A. Chipman, “Polarization Aberrations in Astronomical Telescopes: The Point Spread Function,” *PASP***127**, 445–468 (2015).
11. R. A. Frazin, “Utilization of the Wavefront Sensor and Short-exposure Images for Simultaneous Estimation of Quasi-static Aberration and Exoplanet Intensity,” *ApJ***767**, 21 (2013).
12. J. L. Codona and M. Kenworthy, “Focal Plane Wavefront Sensing Using Residual Adaptive Optics Speckles,” *ApJ***767**, 100 (2013).
13. R. A. Frazin, “Simultaneous ultra-high contrast imaging and determination of time-dependent, non-common path aberrations in the presence of detector noise,” in “Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series,” , vol. 9145 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series (2014), vol. 9145 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series.

[†]“Overloading” is a term borrowed from computer science. An overloaded symbol has two or more meanings depending on the context. For example, the “+” symbol is often overloaded, as it can be used to add numbers, matrices, functions, etc.

14. V. Fathipour, O. G. Memis, S. J. Jang, F. Khalid, R. L. Brown, I. Hassaninia, R. Gelfand, and H. Mohseni, "Isolated nanoinjection photo detectors for high-speed and high-sensitivity single-photon detection," in "Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series," , vol. 8868 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series (2013), vol. 8868 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 3.
15. G. Finger, I. Baker, D. Alvarez, D. Ives, L. Mehrgan, M. Meyer, J. Stegmeier, P. Thorne, and H. J. Weller, "Evaluation and optimization of NIR HgCdTe avalanche photodiode arrays for adaptive optics and interferometry," in "Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series," , vol. 8453 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series (2012), vol. 8453 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series.
16. G. Finger, I. Baker, D. Alvarez, D. Ives, L. Mehrgan, M. Meyer, J. Stegmeier, and H. J. Weller, "SAPHIRA detector for infrared wavefront sensing," in "Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series," , vol. 9148 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series (2014), vol. 9148 of Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, p. 17.
17. B. Mazin, "MKIDs for TMT," in "TMT in the Astronomical Landscape of the 2020s, Thirty Meter Telescope Science Forum, held 16-19 July, 2014 in Tucson Arizona. Online at: <http://conference.ipac.caltech.edu/tmtsf2014/>," (2014).
18. S. Hinkley, B. R. Oppenheimer, R. Soummer, D. Brenner, J. R. Graham, M. D. Perrin, A. Sivaramakrishnan, J. P. Lloyd, L. C. Roberts, Jr., and J. Kuhn, "Speckle Suppression Through Dual Imaging Polarimetry, and a Ground-based Image of the HR 4796A Circumstellar Disk," *ApJ***701**, 804-810 (2009).
19. S. Gladysz, N. Yaitskova, and J. C. Christou, "Statistics of intensity in adaptive-optics images and their usefulness for detection and photometry of exoplanets," *Journal of the Optical Society of America A* **27**, A260000–A75 (2010).
20. A. Labeyrie, "Images of exo-planets obtainable from dark speckles in adaptive telescopes." *A&A***298**, 544 (1995).
21. P. J. Bordé and W. A. Traub, "High-Contrast Imaging from Space: Speckle Nulling in a Low-Aberration Regime," *ApJ***638**, 488–498 (2006).
22. L. Pueyo, J. Kay, N. J. Kasdin, T. Groff, M. McElwain, A. Give'on, and R. Belikov, "Optimal dark hole generation via two deformable mirrors with stroke minimization," *Appl. Opt.* **48**, 6296–6312 (2009).
23. J. W. Goodman, Statistical Optics (John Wiley and Sons, Inc., 1985).
24. D. P. Greenwood, "Bandwidth specification for adaptive optics systems," *J. Opt. Soc. Am.* **67**, 390–393 (1977).
25. M. Born and E. Wolf, Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light, seventh expanded edition (The Press Syndicate of the University of Cambridge, 1999).
26. E. Collett, Polarized Light: Fundamentals and Applications (Marcel Dekker, Inc., 1993).
27. J. P. McGuire and R. A. Chipman, "Diffraction image formation in optical systems with polarization aberrations. i: Formulation and example," *J. Opt. Soc. Am. A* **7**, 1614–1626 (1990).
28. E. Wolf, "Unified theory of coherence and polarization of random electromagnetic beams," *Physics Letters A* **312**, 263–267 (2003).
29. R. A. Horn and C. R. Johnson, Topics in Matrix Analysis (Cambridge University Press, 1994).
30. J. W. Goodman, Introduction to Fourier Optics, second edition (The McGraw-Hill Companies, Inc., 1996).
31. C. Allen, Astrophysical Quantities, third edition (The Athlone Press, LTD., 1991).
32. J. Tervo, T. Setälä, J. Turunen, and A. T. Friberg, "Van cittert-zernike theorem with stokes parameters," *Opt. Lett.* **38**, 2301–2303 (2013).
33. X. Ji and Z. Pu, "Effects of atmospheric turbulence on the polarization of apertured electromagnetic gaussian schell-model beams," *Journal of Optics A: Pure and Applied Optics* **11**, 045701 (2009).
34. F. Roddier, "The effects of atmospheric turbulence in optical astronomy," *Progress in optics* **19**, 281–376 (1981).
35. J. Bailey, Z. Ulanowski, P. W. Lucas, J. H. Hough, E. Hirst, and M. Tamura, "The effect of airborne dust on astronomical polarization measurements," *MNRAS***386**, 1016–1022 (2008).

36. M. S. Olga Korotkova and E. Wolf, “The far-zone behavior of the degree of polarization of electromagnetic beams propagating through atmospheric turbulence,” *Optics Communications* **233**, 225–230 (2004).
37. K. Hinnen, M. Verhaegen, and N. Doelman, “A data-driven h_2 -optimal control approach for adaptive optics,” *Control Systems Technology, IEEE Transactions on* **16**, 381–395 (2008).
38. F. Martinache, O. Guyon, C. Clergeon, and C. Blain, “Speckle Control with a Remapped-Pupil PIAA Coronagraph,” *PASP***124**, 1288–1294 (2012).
39. F. Martinache, O. Guyon, N. Jovanovic, C. Clergeon, G. Singh, T. Kudo, T. Currie, C. Thalmann, M. McElwain, and M. Tamura, “On-Sky Speckle Nulling Demonstration at Small Angular Separation with SCEXAO,” *PASP***126**, 565–572 (2014).
40. T. D. Groff and N. J. Kasdin, “Kalman filtering techniques for focal plane electric field estimation,” *J. Opt. Soc. Am. A* **30**, 128–139 (2013).